ST3123 Introduction of Statistics II

Review

A: Basic Concepts in Inferential Statistics

- Population and Sample(simple random sample)

  Population: ______________________________

  Sample: _________________________________

- Parameter (____________) and statistic(______________)
- Statistics(______________statistics and _______________statistics)
- Inferential statistics (___________________and____________________)
- Central Limit Theorem and sampling distribution of sample mean

1. (CLT) When sample size n is sufficiently **large (_______**), the sampling distribution of sample mean $\bar{x}$ will be approximately a ________ distribution with mean _____________ and standard deviation____________, regardless the population distribution.
The larger the sample size, the better will be the normal approximation to the sampling distribution of $\bar{x}$.

2. When sample size n is small ($n < 30$), and the sample drawn from an approximately normally distributed population with mean μ and population standard deviation σ is known,

3. When sample size n is small ($n < 30$), and the sample drawn from an approximately normally distributed population with mean μ and population standard deviation σ is unknown,

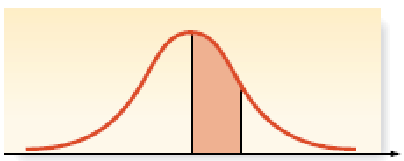
B: two tables

- Standard Normal table (z-table, Table IV, 794)

1. the total area under the z-curve is____;
2. symmetric and bell shaped;
3. the ________z falls in a specific range (z_1, z_2) equals the associated ____under the z-curve.

794 **Appendix A** Tables

TABLE IV Normal Curve Areas



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

Source: Abridged from Table I of A. Hald, *Statistical Tables and Formulas* (New York: Wiley), 1952. Reproduced by permission of A. Hald.

Example1: Given z value, try to find specific area under standard normal curve (p-value).

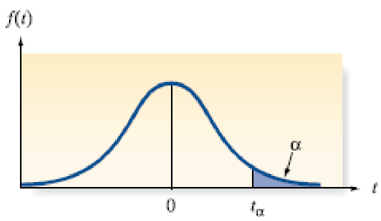
Example2: Given a specific area under standard normal curve, try to find the corresponding z value (critical value).

- **Critical Values of t- table (Table VI, p796)**

1. the total area under the t-curve is_____;
2. symmetric and bell shaped (with heavy tails compared to z-curve);
3. the _________t falls in a specific range (t_1, t_2) equals the associated_____ under the t-curve.
4. Each t-curve is associated with its degree of freedom. When the degree of freedom is getting large, t-curve is close to z-curve.

796 **Appendix A** Tables

TABLE VI Critical Values of *t*



Degrees of Freedom	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$	$t_{.001}$	$t_{.0005}$
1	3.078	6.314	12.706	31.821	63.657	318.31	636.62
2	1.886	2.920	4.303	6.965	9.925	22.326	31.598
3	1.638	2.353	3.182	4.541	5.841	10.213	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.500	2.807	3.485	3.767
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	1.296	1.671	2.000	2.390	2.660	3.232	3.460
120	1.289	1.658	1.980	2.358	2.617	3.160	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.090	3.291

Source: This table is reproduced with the kind permission of the Trustees of Biometrika from E. S. Pearson and H. O. Hartley (eds.), *The Biometrika Tables for Statisticians,* Vol. 1, 3d ed., Biometrika, 1966.

Example3: Given a specific area under t- curve, try to find the corresponding t value (critical value).

Chapter 7&8 Inferences based on a single sample:

Confidence Interval and Tests of hypothesis

- <u>The target parameters for ONE population</u>:

______: population mean

______: population proportion

______: population variance

- Making inferences about the population mean:

1. Chapter7: ____________ with confidence interval

100(1-α)% confidence interval for μ :(large-sample)______________________

(small-sample)_______________________

100(1-α)% confidence interval for p :(large-sample)________________________

2. Chapter8: __________________________

Examples for estimation with confidence interval:

Example1. Unoccupied seats on flights cause airlines to lose revenue**.** Suppose a large airline wants to estimate its average number of unoccupied seats per flight over the past year. 225 flights are randomly selected and results to mean number of unoccupied seats per fight is 11.6, and standard deviation is 4.1. Use 95% confidence interval to estimate μ , the mean number of unoccupied seats per flight during the past year. Interpret the result.

Example2. A random sample of 28 light bulbs had a mean life of 497 hours and a standard deviation of 25 hours. Construct a 90% confidence interval for the mean life μ of all light bulbs of this type. Interpret the result.

Example3. Gun Control, In a survey conducted by Louis Harris of LH research, 1250 U.S. adults were polled regarding their views on banning handgun sales, 650 favored a ban.

1. Use a 95% confidence interval to estimate the proportion of all U.S. adults in favor of banning handgun sales.
2. Based on the confidence interval, can we infer that the majority of U.S. adults in favor of banning handgun sales?

(check if the sample size is large enough)

.

8.1 The Elements of a test of hypothesis

- Elements of a test of hypothesis:

1. **__________________(H_0):** A theory about the values of one or more population parameters. The theory generally represents the status quo, which we adopt until it is proven false.

$$H_0: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_$$

____________________(H_a): A theory that contradicts the null hypothesis. The theory generally represents that which we will accept only when sufficient evidence exists to establish its truth.

2. **Test statistic:** a ________________used to decide whether to reject the null hypothesis.

3. **Rejection region:** The numerical values of the test statistic for which the null hypothesis will be___________. The rejection region is chosen so that the probability of rejecting a true null hypothesis (making a Type I error) is α . α is also called ____________________________.

4. **Conclusion:**

a. If the numerical value of the test statistic ______________**the rejection region**, we __________the null hypothesis and conclude that the alternative hypothesis is true.

b. If the numerical value of the test statistic ________________**the rejection region**, we ________________the null hypothesis and conclude that there is insufficient evidence to conclude that the alternative hypothesis is true.

Note: Conclusions and consequences for a test of hypothesis

	H_0 in fact is	
Conclusion	true	false
Do not reject H_0		
Reject H_0		

The relationship between α and β : ____

When we do the hypothesis test, ____ will be controlled.

Typical values for α are:_______________________**.**

- **Basic logic of hypothesis testing:**

1. Take a ________________ from the population of interest;

2. If the sample data provides ____________________________ to conclude that H_0 **is false**,

   we ________ H_0 and **assert** the alternative hypothesis.

3. If the sample data **does not provide sufficient** evidence to conclude that H_0 **is false**,

   we ______________ H_0 and conclude **there is not enough evidence to assert** the alternative hypothesis.

- **How to set up null hypothesis and alternative hypothesis:**

Three points: 1. What is the __________________?
2. What is the ______________ the parameter will be compared with?
3. What is _______________ between the parameter and the specific value you are interested in comparing?

Examples: Set up null hypothesis and alternative hypothesis; State Type I, II Error

Example1: SSHA: The Survey of Study Habits and Attitudes is a psychological test that measures the motivation, attitude, and study habits of college students. **The mean score is expected 115**. A survey based on 81 incoming freshmen result in mean score is 116.2 and standard deviation is 25. Do the data provide evidence that the mean SSHA score is **different from 115**?

Example2. Gun Control: In a survey conducted by Louis Harris of LH research, 1250 U.S. adults were polled regarding their views on banning handgun sales, 650 favored a ban.
At 10% significance level, do the data provide sufficient evidence to conclude that **more than 50%** of U.S. adults favor banning handgun sales?

8.2 Large-sample test of hypothesis about one population mean

- **Condition required for a valid large-sample hypothesis test for population mean μ :**

1. A ______________is selected from the target population;
2. The sample size n is _______(n >___).

Under these conditions, by Central Limit Theorem,
The Sampling distribution of $\bar{x}$ **is approximately ___________with:**

Mean: $\mu_{\bar{x}} =$

Standard error: $\sigma_{\bar{x}} =$

- Large sample Z-test of hypothesis for a population mean μ :

1. set up null hypothesis and alternative hypothesis

$H_0 :$ ________________________________

$H_a :$ $(two-tailed)$

or $H_a :$ $(lower-tailed)$

or $H_a :$ $(upper-tailed)$

2. significance level α

3. test statistic: $z = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}} \approx$

4. rejection region: _____________when $H_a : \mu \neq \mu_0$

____________ when $H_a : \mu < \mu_0$

______________ when $H_a : \mu > \mu_0$

5. conclusion: if the value of test statistic **falls in** rejection region, ______H_0, and conclude that at α level, there is **sufficient** evidence to conclude H_a is true.

if the value of test statistic **does not fall in** rejection region, _____________H_0, and conclude that at α level, there is **insufficient** evidence to conclude H_a is true.

Examples for Large-sample Z-test of hypothesis about a population mean μ

Example1, SSHA: The Survey of Study Habits and Attitudes is a psychological test that measures the motivation, attitude, and study habits of college students. The mean score is expected 115. A survey based on 81 incoming freshmen result in mean score is 116.2 and standard deviation is 25. Do the data provide evidence that the mean SSHA score is **higher than** 115? Use $\alpha = 0.05$.

Example2. HOSPLOS, The length of stay (in days) in hospital for 100 randomly selected patients are presented in the table. Suppose we want to test the hypothesis that the true mean length of stay (LOS) at the hospital is **less than** 5 days. Use significance level $\alpha = 0.05$.

LOS for 100 hospital patients	2, 3, 8, 6, 4, 4, 6, ………, 10, 2, 4, 2

SPSS output for HOSPLOS,

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
LOS	100	4.53	3.678	.368

One-Sample Test

	Test Value = 5					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
LOS	-1.278	99	.204	-.470	-1.20	.26

8.3 Observed significance levels: p-values

- **p-value(OSL):** the ___________(assuming H_0 is true) of observing a value of the test statistic that is at least contradictory to the null hypothesis, and supportive of the alternative hypothesis, as the actual one computed from the sample data.
- **P-value** is the __________ significance level for which we would reject H_0.

- How to calculate the p-value: (the value of test statistic and H_a)

1. Determine _______________________ based on the sample data;
2. **p-value** = ______________ when $H_a : \mu \neq \mu_0$

   p-value =______________ when $H_a : \mu < \mu_0$

   p-value =______________ when $H_a : \mu > \mu_0$

Examples for p-value calculation:

- How to make test conclusion based on p-value:

If p-value is _____________the significance level α **, ________** H_0**;**

If p-value is _____________the significance level α **, _________________** H_0**.**

Using p-value to make conclusion of Hypothesis test:

Example1, SSHA: A survey based on 81 incoming freshmen result in mean SSHA score is 116.2 and standard deviation is 25. Do the data provide evidence that the mean SSHA score is **higher than 115**? Use significance level $\alpha = 0.05$.

Example2. HOSPLOS, The length of stay (in days) in hospital for 100 randomly selected patients are presented in the table. Do the data provide evidence that the true mean length of stay (LOS) at the hospital is **different from** 5 days. Use $\alpha = 0.05$.

LOS for 100 hospital patients	2, 3, 8, 6, 4, 4, 6, ………, 10, 2, 4, 2

SPSS output for HOSPLOS,

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
LOS	100	4.53	3.678	.368

One-Sample Test

	Test Value = 5					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
LOS	-1.278	99	.204	-.470	-1.20	.26

Example3. HOSPLOS, The length of stay (in days) in hospital for 100 randomly selected patients are presented in the table. Suppose we want to test the hypothesis that the true mean length of stay (LOS) at the hospital is **less than** 5 days. Use $\alpha = 0.05$.

- **How to convert a two-tailed p-value from SPSS output to an one–tailed p-value:**

** 1. **if** $H_a: \mu < \mu_0$ (lower-tailed) and z < 0(negative),

or **if** $H_a: \mu > \mu_0$ (upper-tailed) and z > 0(positive), then $p - value =$______________;

2. **if** $H_a: \mu < \mu_0$ (lower-tailed) and z > 0(positive),

or **if** $H_a: \mu > \mu_0$ (upper-tailed) and z < 0(negative), then $p - value =$__________________.

- Two approaches to do hypothesis test:

A. _________________ Approach:
1. set up null hypothesis and alternative hypothesis,
2. calculate test statistic,
3. find appropriate rejection region,
4. make conclusion. (if test statistic falls in R. R, reject null hypothesis.)

B. ________________ Approach:
1. set up null hypothesis and alternative hypothesis,
2. calculate test statistic,
3. find p-value for the test,
4. make conclusion. (if p-value is smaller than α, reject null hypothesis.)

8.4 Small-Sample test of hypothesis about a population mean

- **Condition required for a valid small-sample hypothesis test for μ:**

1. A ____________________is selected from the target population;
2. The population from which the sample is selected has a distribution that is approximately________________.
3. Sample size is small (   ).

- Small sample t- test of hypothesis for a population mean μ :

1. set up null hypothesis and alternative hypothesis

$$H_0: \ \mu = \mu_0$$

$$\begin{aligned} &H_a: \ \mu \neq \mu_0 \quad (two-tailed) \\ or\ &H_a: \ \mu < \mu_0 \quad (lower-tailed) \\ or\ &H_a: \mu > \mu_0 \quad (upper-tailed) \end{aligned}$$

2. significance level α

3. test statistic: $t =$ with df = ________

4. rejection region: ___________when $H_a : \mu \neq \mu_0$

___________ when $H_a : \mu < \mu_0$

___________when $H_a : \mu > \mu_0$

5. conclusion: if the value of test statistic falls in R.R, reject H_0, and conclude that at α level, there is sufficient evidence to conclude H_a is true.

if the value of test statistic does not fall in R.R, do not reject H_0, and conclude that at α level, there is insufficient evidence to conclude H_a is true.

- How to find the p-value: (the value of test statistic and H_a**)**

1. Determine **the value of test statistic** $t_0 = \dfrac{\overline{x} - \mu_0}{s/\sqrt{n}}$ based on the sample data;

2. **p-value** = _____________ when $H_a : \mu \neq \mu_0$

   p-value =_______________ when $H_a : \mu < \mu_0$

   p-value =_______________ when $H_a : \mu > \mu_0$

Note: since t-table in our textbook doesn't give the probability, we can't use t-table to find exact p-value, we can only approximate the p-value, but we can use statistical software or SPSS output to get the exact p-value.

Examples for Small-sample t-test of hypothesis about a population mean μ

Example1. EMISSION,

A major car manufacturer wants to test a new engine to determine whether it meets new air-pollution standards. The mean emission of all engines of this type must be **less than 20** parts per million of carbon. Ten engines are manufactured for testing purposes. The data is listed below.

Emissions	15.6	16.2	22.5	20.5	16.4	19.4	19.6	17.9	12.7	14.9

Do the data supply sufficient evidence to allow the manufacturer to conclude that the type of engine **meets the pollution standard**? Assume to risk a type I error with probability α = 0.01.

SPSS output for EMISSION,

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Emission	10	17.5700	2.95223	.93358

One-Sample Test

	Test Value = 20					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Emission	-2.603	9	.029	-2.43000	-4.5419	-.3181

Example2. Mongolian desert ants,

To study the ants in Mongolia, the botanists placed seed baits at 11 sites and observed the number of ant species attracted to each site. Do the data indicate that the average number of ant species at Mongolian desert sites is **greater than 5 species**? Use $\alpha = 0.10$.

# of ant species	3, 3, 52, 7, 5, 49, 5, 4, 4, 5, 4

SPSS output for Example2. Mongolian desert ants,

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
ANTS	11	12.82	18.675	5.631

One-Sample Test

	Test Value = 5					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
ANTS	1.388	10	.195	7.818	-4.73	20.36

8.5 Large-sample Test of hypothesis about a population proportion p

- **Condition required for a valid large-sample hypothesis test for p :**

1. A____________ sample is elected from a binomial population;
2. The sample size n is _____(This condition will be satisfied if ______________________)

- Large-sample test for a population proportion p :

1. set up null hypothesis and alternative hypothesis

$$H_0 :$$

$$\begin{aligned} H_a &: \quad (two-tailed) \\ or\ H_a &: \quad (lower-tailed) \\ or\ H_a &: \quad (upper-tailed) \end{aligned}$$

2. significance level α

3. test statistic:

4. rejection region: ______________when $H_a : p \neq p_0$

_____________ when $H_a : p < p_0$

______________when $H_a : p > p_0$

5. conclusion: if the value of test statistic falls in R.R, reject H_0 , and conclude that at α level, there is sufficient evidence to conclude H_a is true.

if the value of test statistic does not fall in R.R, do not reject H_0, and conclude that at α level, there is insufficient evidence to conclude H_a is true.

Examples for Large-sample test for a population proportion p :

Example1. Gun Control: In a survey conducted by Louis Harris of LH research, 1250 U.S. adults were polled regarding their views on banning handgun sales, 650 favored a ban.
At 10% significance level, do the data provide sufficient evidence to conclude that **more than half** of U.S. adults favor banning handgun sales?

Example2. Shipment Defectives,
The reputations of many businesses can be severely damaged by shipments of manufactured items that contain a large percentage of defectives. A manufacturer of alkaline batteries wants to be reasonably certain that fewer than 5% of its batteries are defective. Suppose 300 batteries are randomly selected from a large shipment, 10 defective batteries are found.
Does this provide sufficient evidence for the manufacturer to conclude that **the fraction** defective in the entire shipment is **less than 0.05**? Use α = 0.01.